学校编码：10384　　分类号_____密级_____

学号：X2008230097　　UDC _____

厦门大学

硕 士 学 位 论 文

企业邮件防泄密审计系统的设计与实现

Design and Implementation of Enterprise E-mail Anti-phishing Audit Systems

邱 贵 强

指导教师姓名：史 亮 副教授

专 业 名 称：软 件 工 程

论文提交日期：2010 年 月

论文答辩日期：2010 年 月

学位授予日期：2010 年 月

答辩委员会主席：_____

评 阅 人：_____

2010 年 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下，独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果，均在文中以适当方式明确标明，并符合法律规范和《厦门大学研究生学术活动规范（试行）》。

另外，该学位论文为（　　）课题（组）的研究成果，获得（　　）课题（组）经费或实验室的资助，在（　　）实验室完成。（请在以上括号内填写课题或课题组负责人或实验室名称，未有此项声明内容的，可以不作特别声明。）

声明人（签名）：

年　　月　　日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（　　）1.经厦门大学保密委员会审查核定的保密学位论文，于　　年　　月　　日解密，解密后适用上述授权。

（　　）2.不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年　　月　　日

摘　要

随着计算机技术和网络技术的不断发展，电子邮件的应用逐渐深入人们的工作生活当中，已经成为现代通信不可或缺的一部分。由于电子邮件自身的特点以及部分互联网用户缺少相应的保密意识，近年来通过电子邮件的窃密泄密行为越来越多，对企业邮件防泄密安全审计的问题摆在科研人员的面前。

论文首先研究了近年来国内外对电子邮件内容审计研究技术的发展现状及前景，还研究了电子邮件系统的构成、电子邮件报文结构、数据库技术等。其次作者还根据用户的业务需求，结合作者本人参与的项目研发工作，将邮件监听系统与数据库技术进行融合，提出了基于旁路监听的企业电子邮件防泄密安全审计系统总体结构的建设。由于对邮件防泄密审计不需要系统对样本进行自行学习，论文还对邮件预处理算法进行改进，简化一些不必要的算法，提高了系统对数据的处理能力，对关键词匹配技术做了优化，最后根据业务需求设计了后端处理系统，提高系统处理电子邮件数据的准确性及智能化。

关键字：电子邮件;旁路监听;安全审计

Abstract

With the rapid development of computer technology and network technology, e-mail has been very popular in people's life and plays an important role in modern communication. In recent years, because of the features email employ as well as the lack of appropriate sense of confidentiality for Internet users, disclosure and acts of theft by e-mail turn out to be a very serious problem in many enterprises, it is very urgent and necessary for the researchers to solve this problem by implementing the anti-phishing security auditing .

This dissertation makes the study on the development of research technology for e-mail content auditing at home and abroad in recent years, and further gets insights to the composition of the e-mail system, e-mail message structure and database technology. Furthermore, based on the user's actual requirements and the experiences generated from the involvement in research and development, the author also provides the scheme for establishing anti-phishing security auditing system by incorporating bypass monitoring. Since the anti-phishing auditing does not need to adopt the anti-phishing self-learning system for samples, this paper also offer the suggestions for improving the email preprocessing algorithm ,aiming to simplify the unnecessary algorithm so as to improve the system's capability in data processing and optimizing of keyword matching technology .Finally, according to the business needs , the back-end processing system is designed to improve the accuracy and intelligence of the email processing for data system .

Keywords**:** E-mail ;bypass monitoring;security audit

目录

Table of Contents

第一章 绪论

1.1 研究背景

电子邮件诞生于上个世纪70年代初，由于技术与网络带宽的限制，当时只能发送简短的信息，电子邮件只是在研究所的科研人员中得到应用和发展。1987年9月20日，北京计算机应用技术研究所通过一条电话线发送第一封电子邮件给德国卡尔斯鲁厄大学，揭开了中国使用电子邮件的序幕。随着互联网技术的高速发展，电子邮件从科研教育单位开始向公众普及，它凭借着远远优于传统邮政系统的收发迅速、费用低廉以及使用方便等优势，逐渐成为人们相互交流、获取信息的重要渠道。中国互联网信息中心（CNNIC）2010年7月发布的《CNNIC第26次中国互联网发展状况调查报告》指出，2010年上半年，我国网民继续保持增长态势，截至2010年6月，总体网民规模达到4.2亿，突破了4亿关口，较2009年底增加3600万人。互联网普及率攀升至31.8%，较2009年底提高2.9个百分点。在电子邮件应用方面，2009年12月的使用率为56.8%，在所有调查的应用中排名第8位，2010年6月的使用率为56.5%，在所有调查的应用中排名第7位[1]。

随着电子邮件技术及其相关安全技术的成熟，各行各业已逐渐将电子邮件应用到了一些正式场合，例如使用电子邮件来传递大量社会、经济、政治等重要信息。由于电子邮件系统的自身存在的缺陷以及使用者缺少必要的安全防范意识，可能为有意或无意的泄密行为带来便利，统计数据表明，相当多的安全事件是由内网用户有意或无意的误用造成的[2, 3]。其次电子邮件泄密信息的内容会被以最快的速度传递给对方，并且一般人丝毫不会察觉，所以电子邮件泄密，隐蔽性更强、危害性会更大。近年来随着电子邮件泄密事件频频被揭露，企业级电子邮件安全及重要电子文档防泄密等相关重大网络安全问题已经引起了社会各界的广泛重视。近年来信息安全专家一直致力于信息安全方面的研究，并取得了一定的成果，比如防火墙技术，防火墙技术可以在一定程度上防止外来的入侵，却没法防止企业内部向外因缺乏安全意识或有意的泄密、窃密行为。

目前对于企业内网邮件防泄密内容审计有两大类：一种发现违规邮件阻断发送，一种是基于旁路监听的方式对邮件进行审计，前者可对违规邮件进行预

防性的阻断，但其会因误报而引起部分邮件使用者的抵触情绪，对于有意泄密者却能进一步进行规避；后者不影响邮件的使用，主要用于泄密线索的追查和搜集证据。本论文的研究课题就在企业内网信息安全范围内，针对邮件泄密领域对邮件进行旁路监听审计的信息安全系统的设计与实现，特别是对后端应用方面进行研究与实现。

1.2 国内外的研究发展现状

1.2.1 国内外邮件审计的方式

目前国内外常用的电子邮件审计可分为邮件服务器文件扫描和基于旁路监听两种方式。

一、基于邮件服务器文件扫描系统的电子邮件审计系统，主要扫描邮件服务器中的邮件，对其内容进行审计。

该方式的主要优点：

1.系统部署于邮件服务器本地，可以通过直接访问物理存储器对需要审计的邮件进行访问，速度快，且需要增加投入的硬件设备少；

2.电子邮件服务器的电子邮件结构完整，可以直接读取邮件的内容，不需要进行重组。

其不足之处在于当邮件服务器不在审计人员的管辖范围之内，就无法应用这种方式进行审计，如电子邮件服务器在境外，或者企业本身正从事非法活动等。

二、基于旁路监听的电子邮件审计系统通过对旁路监听到的数据包进行重组，把传输层的数据恢复到应用层，对邮件的内容进行审计[4]。

该方式的优点在于：部署比较稳蔽，适用于对泄密行为进行证据收集。

不足之处在于不能实时阻断泄密行为的发生[5]，硬件投入比较庞大，如企业有多个路由出口，需要进行分布式监听。

1.2.2 国内外在基础研究方面的发展现状

一、国外学者在邮件过滤的基础理论研究和算法研究方面起步较早。在关

键词匹配技术方面，国外学者提出了以下几种算法：

1.朴素匹配算法 这种算法要求在正文串中匹配是否包含子串（又称模式串），把正文串与子串对齐后每个字符一一匹配，如果不匹配，子串后移一位与正文错开一位，重新一一进行匹配，这是一种穷举式的匹配模式，效率较为低下，时间复杂度比较大。

2.KMP 算法 这是由 Knuth、Morris、Pratt 提出一种对朴素匹配算法的改进算法，它利用了朴素匹配算法中匹配过的结果，跳过一些不必要比较的字符，使得时间复杂度大大减小。

3.BM 算法　BM 算法是 1977 年 Boyer 和 Moore 提出的一种与 KMP 类似的算法，其特点是字符串从后向前匹配，并提出了坏字符与好后缀的概念，其匹配速度比 KMP 快 3－5 倍。

二、在基于内容的垃圾邮件过滤技术方面，国外学者提出了基于规则和基于统计的方法。

基于规则的方法有：

1.决策树方法 实质上是从训练集中学习，得到以决策树的形式表示的分类规则；

2.Boosting 方法 通过弱规则进行加权求和得到强规则的一种方法;

3.Ripper 方法 由 Cohen 提出的一种基于规则的方法，比传统的规则方法速度更快、性能更高；

4．粗糙集方法 由 Pawlak 提出的一种研究不完整、不确定知识和数据的表达、学习、归纳的理论方法。

基于统计的算法有：

1.kNN 方法 一种基于实例的常用方法，没有训练过程，直接将待分类样本与训练集中的样本进行比较；

2.SVM 通过构造最优线性分类面来指导分类，可以直接用于线性可分问题；

3.叶贝斯方法 Thomas Bayes 提出的一种基于概率的算法，被应用于邮件样本的学习，具有智能学习的能力；

4．Rocchio 方法 是信息检索领域常常用于相关反馈的方法。

三、国内的学者在算法方面较少提出一些原创性的算法，主要研究工作是对现有的一些算法进行改进。在中文邮件处理方面,国内学者提出的较有特色的中文分词技术。由于中文邮件与英文邮件词语具有的不同特点，英文中的单词以空隔进行分隔，中文的词是由先后顺序组合一起的汉字通过语义组合成的，所以处理中文信息时中文分词是一项至关重要的环节，国内学者在中文分词方法的研究方面具有先天的优势。现有的邮件中文分词技术主要有：

1.基于词典的分词方法 这种方法是将中文信息与汉语电子词库进行匹配，匹配成功则标识成一个词，这种算法的准确性依靠电子词库的的词数，执行效率不太高且会产生歧义字段，最常用的是最大匹配算法（MM 法）。

2.基于理解的分词技术 这种分词方法通过模拟人对句子的理解，达到识别词的效果，这种算法目前尚处于试验阶段。

3.基于统计的的分词方法 计算相邻字出现的频率，计算它们的互信息，互信息体现汉字间的紧密程度，当互信息达到一定的阈值，则认定为词。常见的方法有：基于 EM 肯德基法的分词、基于 N-gram 模型等。

1.2.3 国内外在应用方面的发展现状

从公开资料了解到美国利用在互联网的资源和技术上的存备，在互联网上建立庞大的监听系统。如：1998 年美国联邦调查局（FBI）为侦破“毒品走私”等组织犯罪开发了“Carnivore（食肉猛兽）”（后改名为“DCS1000”），其公开的软件基本功能有：

一、监听可疑电子邮件的头文件或全部内容；

二、列出服务器可疑的访问（入侵）；

三、全面嗅探可疑的 IP 地址；

四、通过 RADIUS 登陆发现正在网上的可疑 IP 地址；

另外，美国国家安全局（NSA）的“阶梯计划”可对全球传递的电子邮件、传真、电话等进行监听。

近年来国内很多公司对邮件的安全问题都投入了大量的人力、物力进行研究开发，比如金笛电子邮件网关、亿邮邮件网关等，但这些公司主要研究垃圾邮件的过滤等。对于旁路监听的电子邮件审计系统，目前国内有几家公司、研